Visualization Techniques for Intrusion Detection

William Wright and Peter Clarke
Point of Contact: William Wright
Oculus Info Inc.
572 King Street West, Suite 200
Toronto, Ontario M5V 1M3
CANADA

bill.wright@oculusinfo.com

INTRODUCTION

This paper reports on the experiences of using interactive animated 2D and 3D graphics in an Intrusion Detection (ID) Analysts Workbench prototype. Visualization techniques allow people to see and comprehend large amounts of complex data. Graphics are used to assist with the ID investigation and reporting process by helping the analyst identify significant incidents and reduce false conditions (positives, negatives and alarms). Visualization is then used in reporting incidents to a broader senior level audience. Complex patterns are clearly displayed over time in an easy to understand and compelling manner. Initial evaluations of the prototype have been positive, and a second development stage has been initiated.

ID ISSUES

Large numbers of events are generated by network intrusion detection sensors; however not all these events are malicious in nature, not all malicious events are applicable to a given network environment and, perhaps of even more concern, certain malicious events can be missed.

There are several emerging trends in enterprise networking that are making traditional signature based intrusion detection more challenging. The increase use of very high-speed lines and more prevalent use of encryption technology are a challenge for the intrusion detection community. As the data collected becomes larger in volume, or the increasing dependence on traffic pattern anomaly detection as a workaround for payload encryption becomes more widespread, the amount of data the analyst must cope with increases.

Through the use of various types of detection tools and techniques, including signature based network intrusion detection, anomaly based network intrusion detection, and full packet capture, a better picture can be formed. The analyst is able to fuse this data and gain a more comprehensive insight into what is truly of malicious nature.

The massive amounts of data involved in this type of thorough multi-source analysis make it infeasible for most organizations. The significance of the events contained within the data can often only be determined by scanning the huge amounts of data looking for subtle and sometimes unexpected patterns and correlations.

This investigative process is required in order to place the events around an alarm in context and to assess if further action is required, but the process is labor intensive. Fused logs for a short period can easily contain tens to hundreds of thousands of records. Tools are needed to help accomplish this investigative task in less time.

Paper presented at the RTO IST Workshop on "Massive Military Data Fusion and Visualisation: Users Talk with Developers", held in Halden, Norway, 10-13 September 2002, and published in RTO-MP-105.

Report Documentation Page

Form Approved
OMB No. 0704-0188

Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
00 APR 2004	**N/A**	**-**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
Visualization Techniques for Intrusion Detection	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Point of Contact: William Wright Oculus Info Inc. 572 King Street West, Suite 200 Toronto, Ontario M5V 1M3 CANADA**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release, distribution unlimited

13. SUPPLEMENTARY NOTES
See also ADM001665, RTO-MP-105 Massive Military Data Fusion and Visualization: Users Talk with Developers., The original document contains color images.

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			
unclassified	**unclassified**	**unclassified**	**UU**	**26**	

Another issue is that network ID sensors are not always effective for detecting new exploits or for activities that span many weeks and / or multiple network systems. To detect and investigate these types of activities requires the analyst to review extremely large amounts of packet level data.

A related problem is in reporting the attacks and the nature of those attacks to senior managers. This is important in order to raise awareness and provide an understanding of the need for information technology security in industry and government. Without this senior level understanding and support, obtaining security funding can be challenging. The output of most ID processes can be cryptic, and inaccessible to non-experts.

SOLUTIONS

Two graphical consoles have been built to evaluate the usefulness of visualizing intrusion data. Figure 1 shows the Intrusion Detection Analysts Workbench. Up to 2,000,000 event records or more can be displayed and analyzed in multiple concurrent dynamic charts. Each event record includes fields such as source and destination IP, port ID, alarm code, date, time. The charts are scaleable so that, for example, a bar chart showing number of events by destination IPs can easily display ten's of thousands of IP addresses. The charts are also linked. Selecting events in one chart will highlight those events in all the other charts. So for example, selecting events associated with one type of alarm will cross reference those events in the source IP bar chart, and the destination IP bar chart. The analyst workbench is used to investigate, isolate and prioritize events. It was evaluated in a side-by-side test with existing methods and proved to be a significantly faster method. The workbench makes use of the commercial off-the-shelf Advizor product.

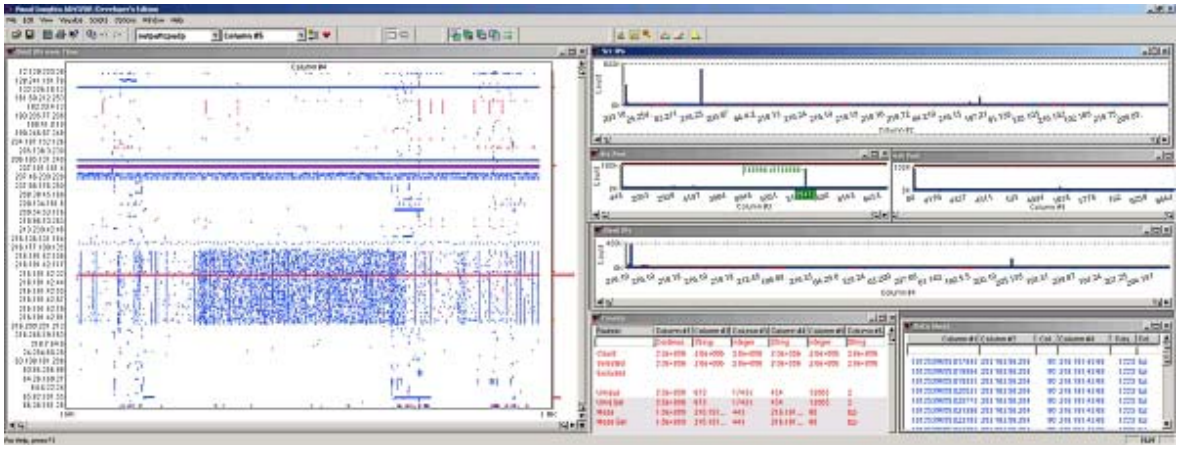


Figure 1: Intrusion Detection Workbench with 2M TCP and UDP Records.

The Analysts Workbench graphical tool can concurrently display the raw packet or alarm data as well as output from analytical tools that, for example, filter events or compute statistical metrics.

Figure 2 shows the Animated Incident Reporting component. It is used to report intrusion activity to senior management, and is designed to show the significance and nature of the events without overwhelming the viewer. The objective is to clearly see who did what to whom and when. A number of interactions are supported including filtering and an adjustable playback speed. This component was evaluated in a series of presentations to senior levels of government and industry.

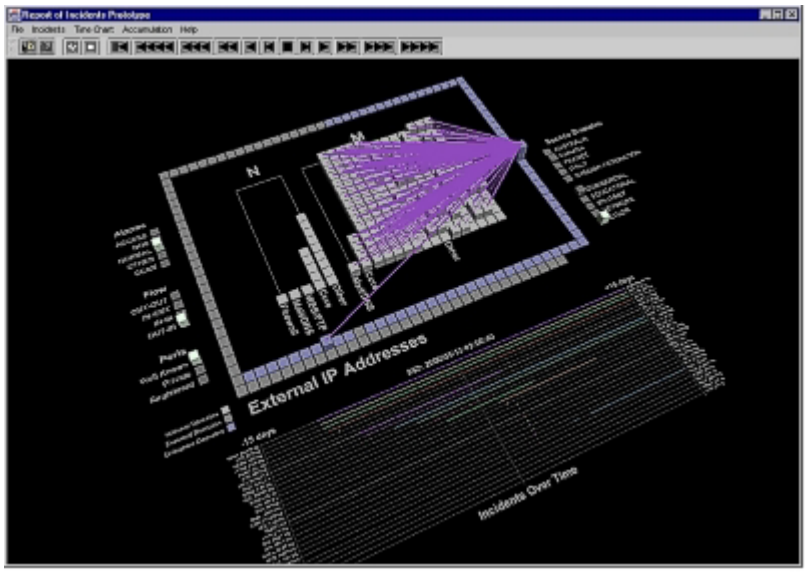


Figure 2: Animated Incident Reporting.

FUTURE DEVELOPMENT

Future work involves two separate but related streams:

The first is the expansion and integration of the two visualization tools to create a seamless intrusion detection visualization workflow environment. Given that intrusion detection analysis is often only part of a systems administration function, time is a consideration. The more effectively the visualization tools can be adapted to fit, and enhance, the human decision making process (orient, observe, decide and act), the more incidents can be effectively assessed and escalated or discarded in a shorter time period.

The second is work on migrating the tools towards an anomaly detection capability through the use of raw network data along with the fused intrusion detection alarms to gain a more comprehensive view into the network.

CONCLUSION

People excel in detecting patterns and identifying relationships when data is presented visually. Extremely large amounts of data can be viewed and compared. This is a useful ability for the ID analyst.

Experience with the visualization methods used in this work has lead to observations and recommendations for developing new methods.

Initial evaluations of the prototype have been positive, and a second development stage has been initiated. The objectives of this second stage will be discussed in the paper.

SYMPOSIA DISCUSSION – PAPER NO: 15

Author's Name:

Mr. William Wright, Oculus Info Inc, Canada
Presented by Ms. Pascale Proulx, Oculus Info Inc, Canada

Question:
What type of methodology was applied in the design of the visualisation?

Author's Response:
User consultations.

Comment:
The system is ten times faster than the previous system that did not use any visual display at all.

Comment:
These display techniques would be useful for all statistical data such as traffic jams.

Question:
Is it possible to transfer the knowledge of the operator into rules for the system to automate the process?

Author's Response:
It may be possible to recognize the patterns, but it is important to investigate more to see what is generating the pattern. For example, a pattern that initially appears dangerous may only be a virus definition update.

Question:
In advance of developing this system, was there consideration of mathematical methods that maybe amenable to clustering and statistical analysis?

Author's Response:
There is a trade off between the math and the visual analysis, as well as between implementation time of the math and algorithms and human interaction. There is research going on in this area right now.

Comment:
Change detection might be important to an analyst.

Question:
There are clustering techniques that could be applied, but require a large amount of training data to make the systems work well. How much data is available for training?

Author's Response:
There currently are not many full data sets. There is a conference in LA that puts hackers against professions, but that produces a data set that is not necessarily typical. Also, networks are dynamic, so constant retraining is required.

Visualization Techniques for Intrusion Detection

William Wright, Sr. Partner, Oculus, Canada
bill.wright@oculusinfo.com
and
Peter Clarke, DND, Canada

Presented by
Pascale Proulx, Visualization Consultant, Oculus
pascale.proulx@oculusinfo.com

Paper first given at: Workshop on Statistical and Machine Learning Techniques in Computer Intrusion Detection, Johns Hopkins University, June 11-13, 2002.

What is Covered?

2

- Intrusion detection issues
- Using visualization as a solution
- Current visualization tools developed
- Future development of visualization in intrusion detection

Confidential and © 2002 Oculus Info Inc. and Others

Intrusion Detection Issues

3

- Large amounts of IDS (Intrusion Detection Sensor) data
  - 1 gigabyte of information will fill a pickup truck with printed paper
- Bad signal/noise ratio on most un-tuned IDS
  - Worse than TV filled with “snow”
- If alarms are removed, harmful events may slip through unnoticed
- Event Correlation (IDS, routers, firewalls)
  - Very important to gaining a complete picture, but makes data handling even more difficult

Confidential and © 2002 Oculus Info Inc. and Others

Intrusion Detection Issues

4

- Reporting incidents to senior management or other non-experts
- The problems are getting worse as technology progresses and network speeds (i.e. bandwidth) increase

Confidential and © 2002 Oculus Info Inc. and Others

Visualization as a Solution

5

- Visualization allows people to see and comprehend large amounts of complex data in a short period of time
- Helps the analyst to identify significant incidents and reduce time wasted with false positives
- Report incidents to a broader, non-expert audience
- Ability to cue the analyst through the use of colour, shape, patterns, or motion

Confidential and © 2002 Oculus Info Inc. and Others

Visualization Tool Development

6

- Two graphical applications have been built for evaluation
  - Intrusion Detection Analyst Workbench
  - Animated Incident Explanation Engine
- Each displays data visually, but currently have two separate audiences

Confidential and © 2002 Oculus Info Inc. and Others

Intrusion Detection Analyst Workbench

7

- More than 2 million events can be displayed and analyzed in multiple concurrent dynamic charts
- Each chart is linked, allowing the analyst to select something in one chart, and it will highlight the relevant details in the other charts
- High performance interactive analysis possible for 2M records on high end Windows / Intel machine – 2 GHz, 1 GB RAM and good graphics card.

Confidential and © 2002 Oculus Info Inc. and Others

Intrusion Detection Analyst Workbench

8

- Assists in isolating, investigating and prioritizing events
- Evaluated side-by-side with existing methods and proved to be significantly faster and easier
- Run by commercial off-the-shelf Advizor™ product

Confidential and © 2002 Oculus Info Inc. and Others

Log Data Fields

9

- Date
- Time
- Direction — out-in, in-out, out-out, in-in
- Alarm Code — (alarms and/or normal traffic)
- Alarm Name
- Source IP
- Destination IP
- Port ID - Source
- Port ID - Destination
- Port Name — ~10 well known, 65,000 possible
- Sensor ID

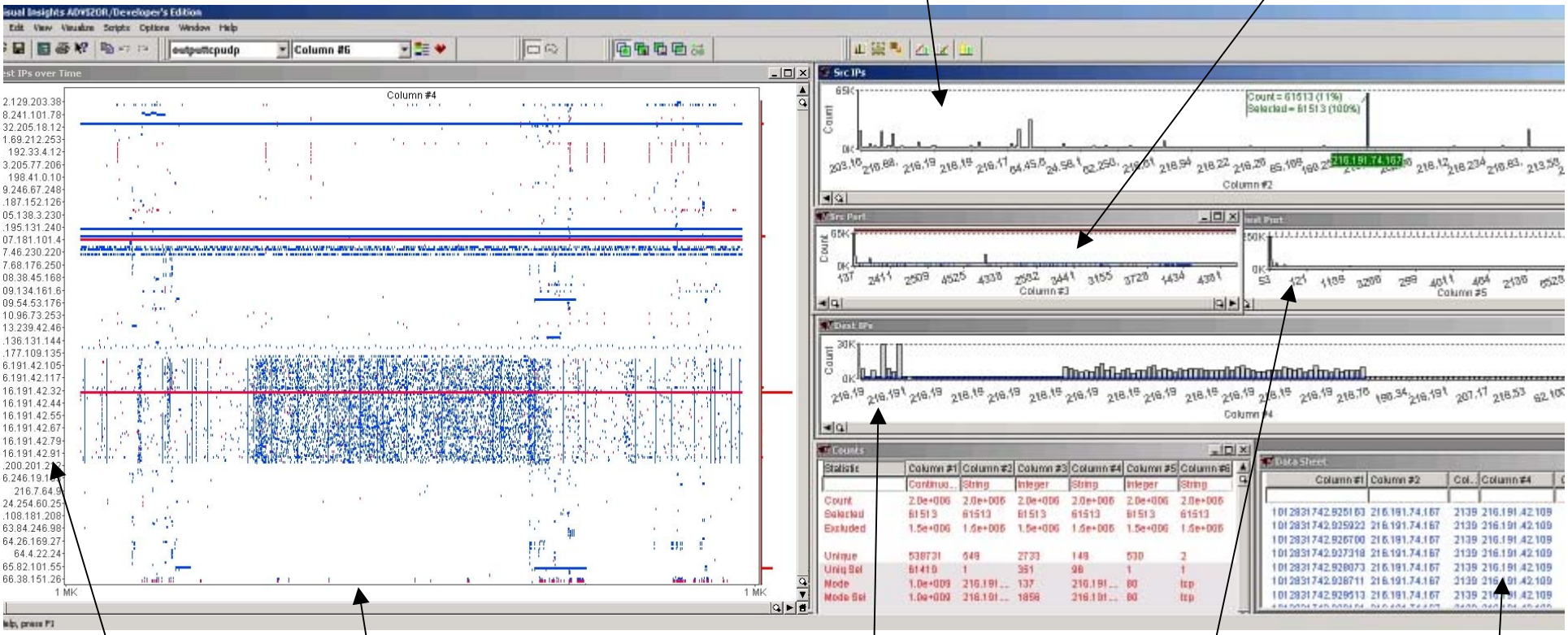
Confidential and © 2002 Oculus Info Inc. and Others

Intrusion Detection Analysts Workbench

Layout consists of multiple charts. Each chart has all the records.

10

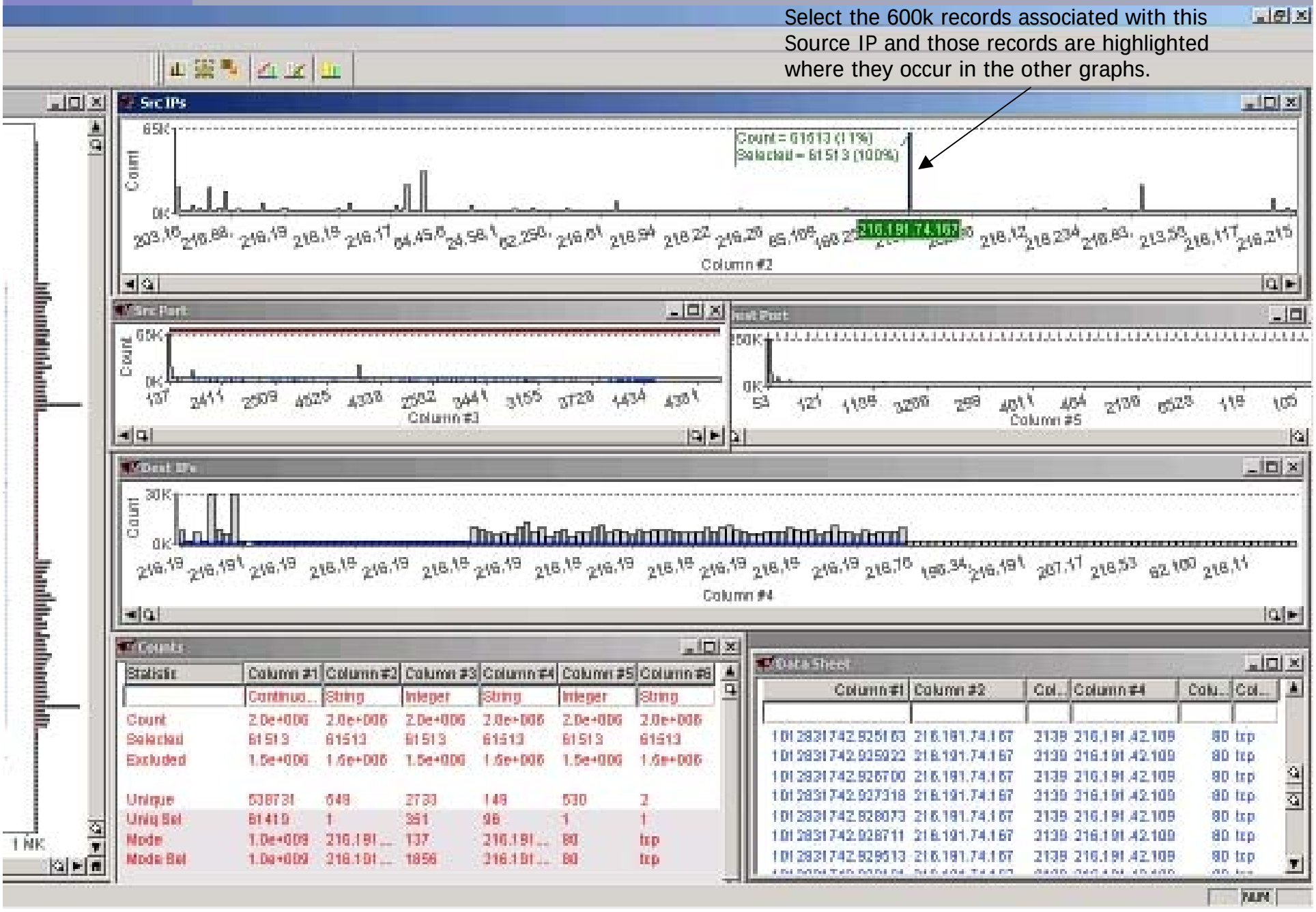
All TCP UDP – 2M records in each chart
(And that's only for a two person network for five days!!)



Confidential and © 2002 Oculus Info Inc. and Others

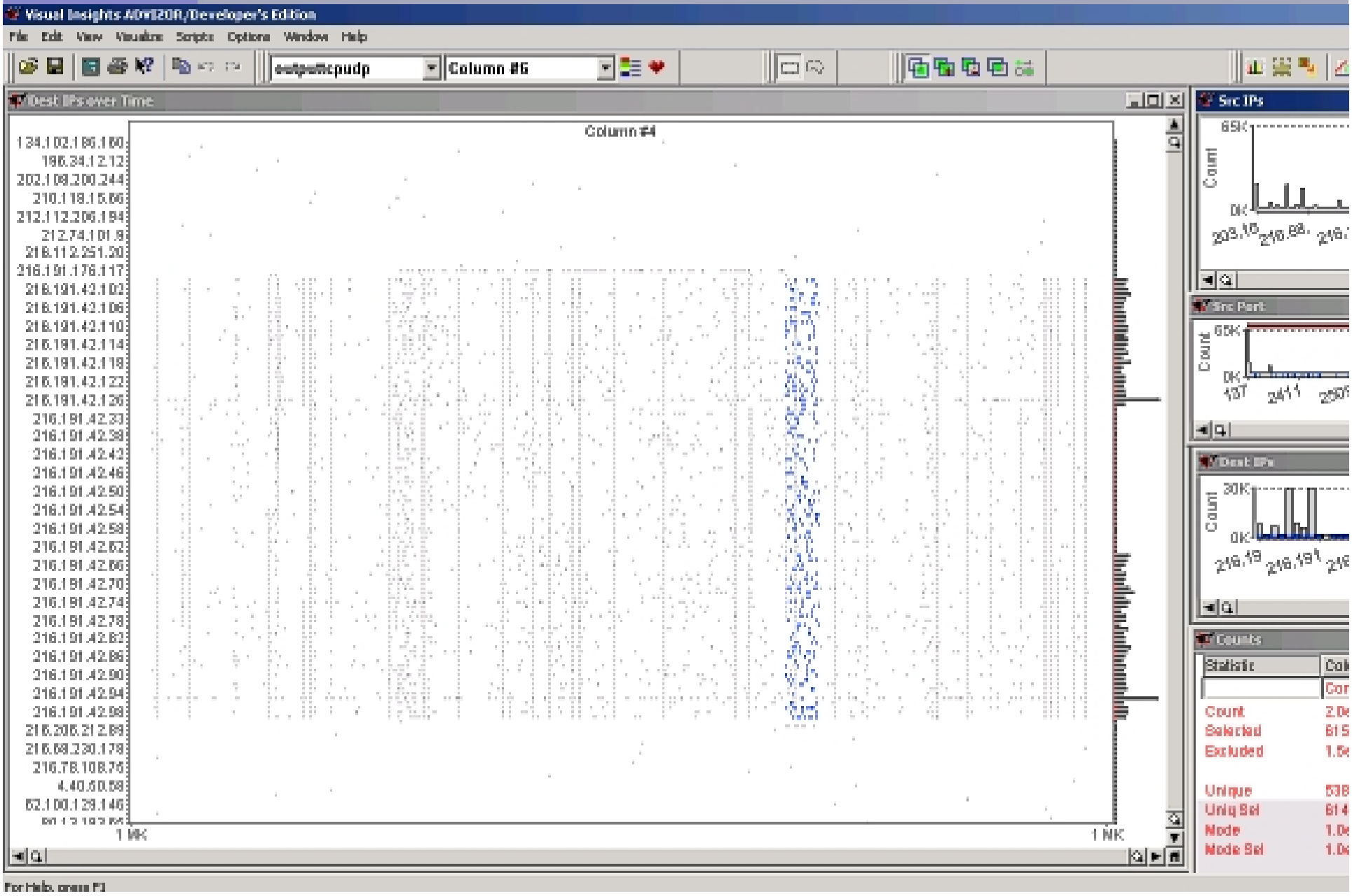
Linked Charting

Select the 600k records associated with this Source IP and those records are highlighted where they occur in the other graphs.



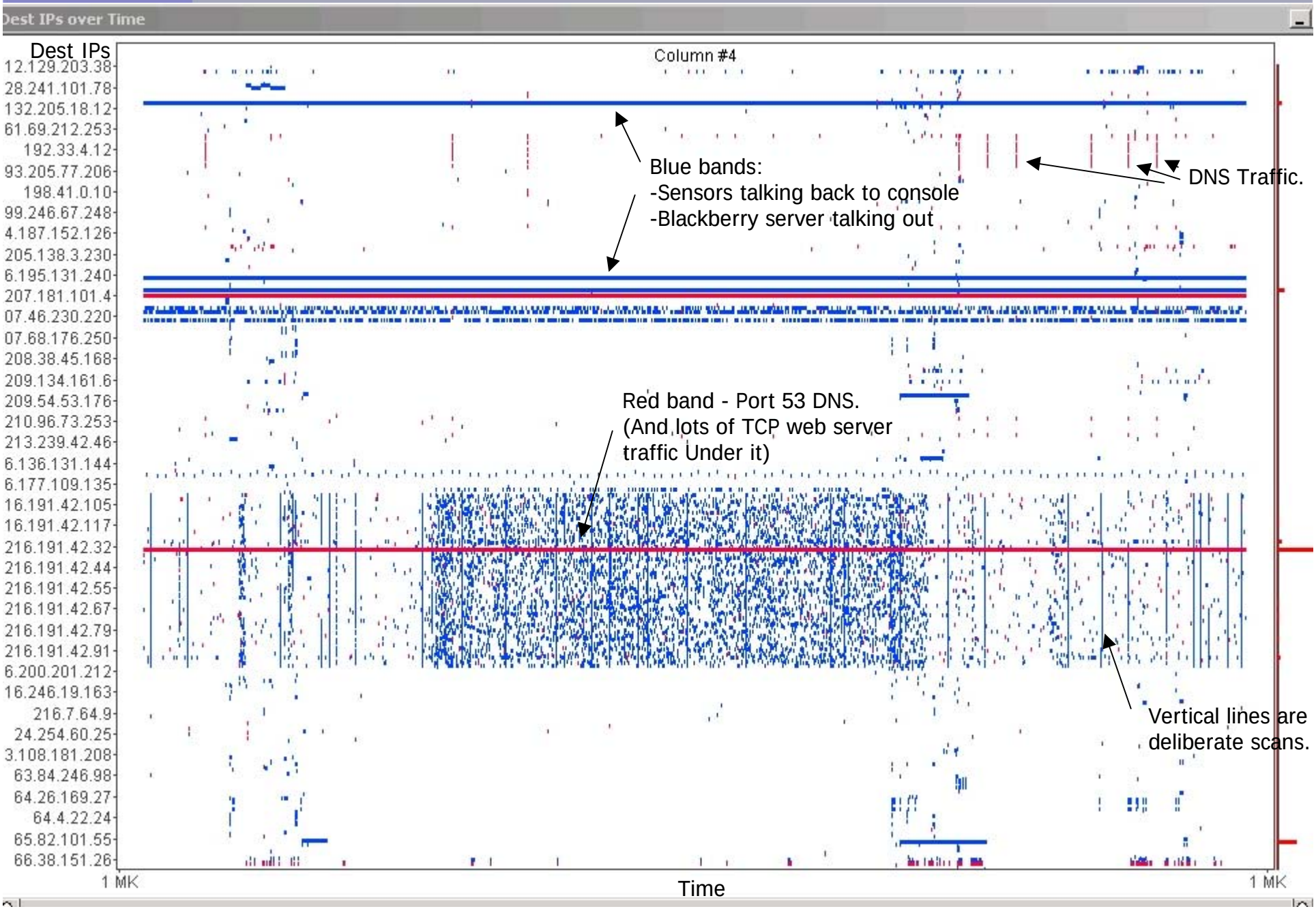
oculus

Linked Charting



Analysis - All TCP UDP – 2M Records

TCP is Blue and UDP is Red. Five days. Two people.



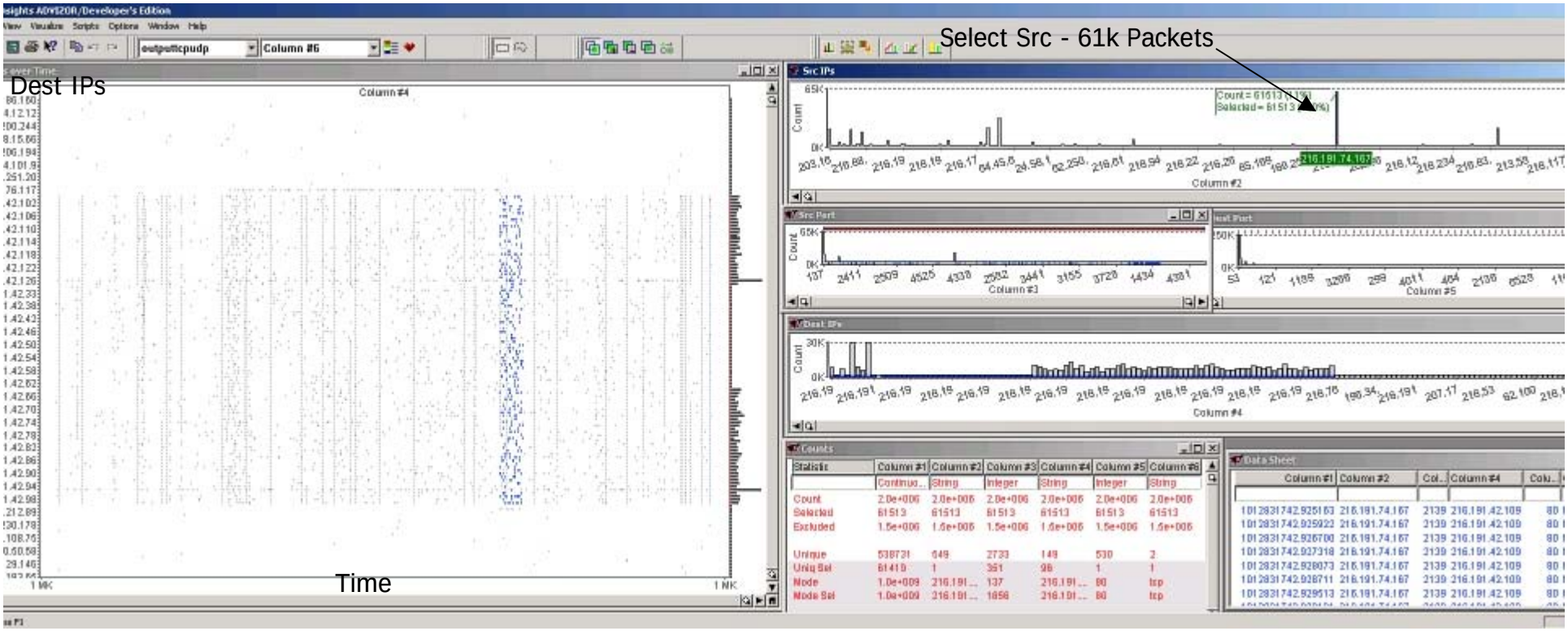
Tests

2M TCP and UDP Records

14

1. More characteristics made visible.

2. Scatter Scan



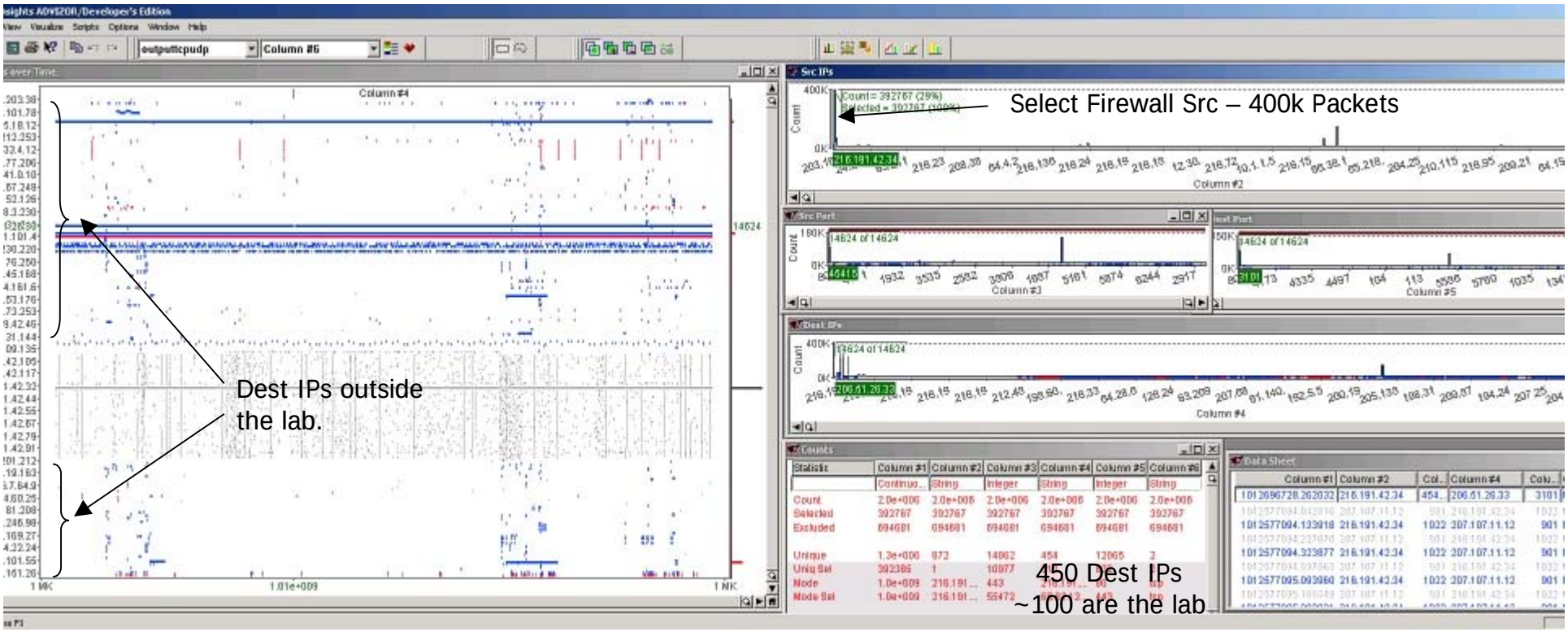
Confidential and © 2002 Oculus Info Inc. and Others

Tests

2M TCP and UDP Records

15

Traffic Outbound – Firewall Src



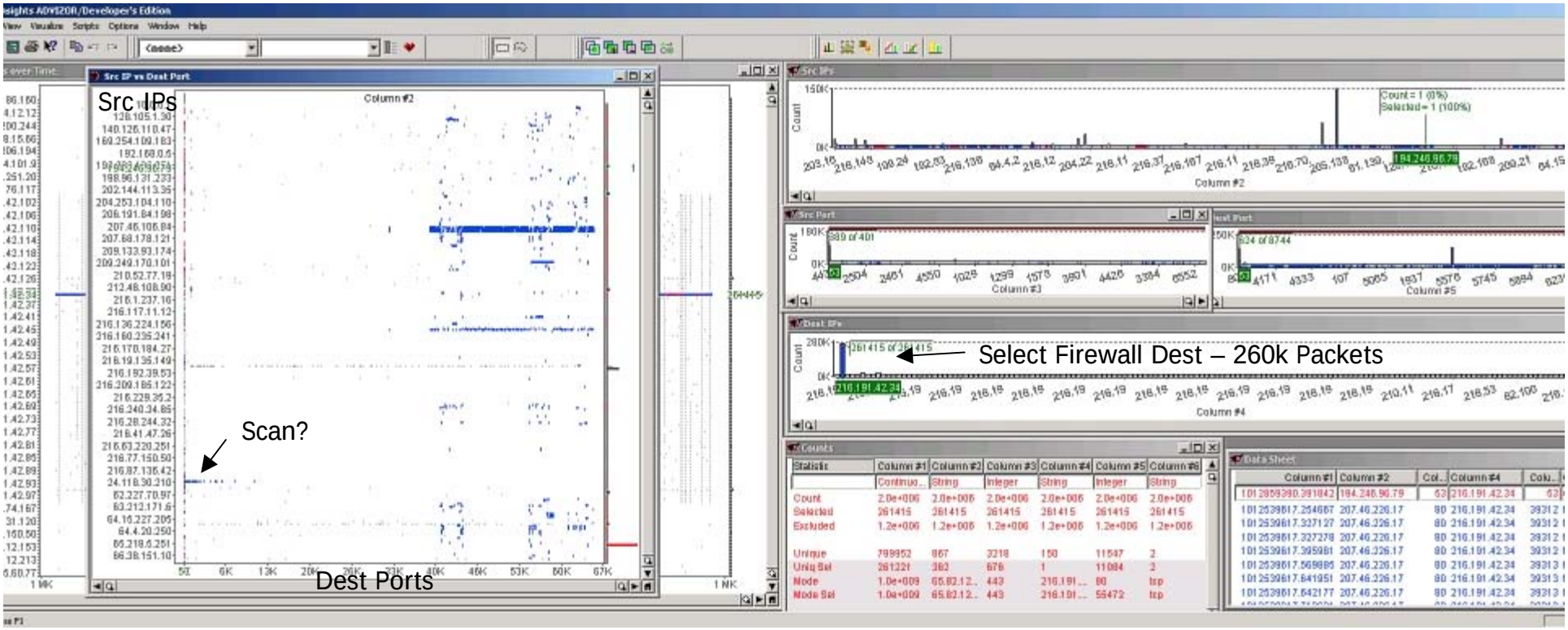
Confidential and © 2002 Oculus Info Inc. and Others

Tests

2M TCP and UDP Records

16

Low Port Scan?



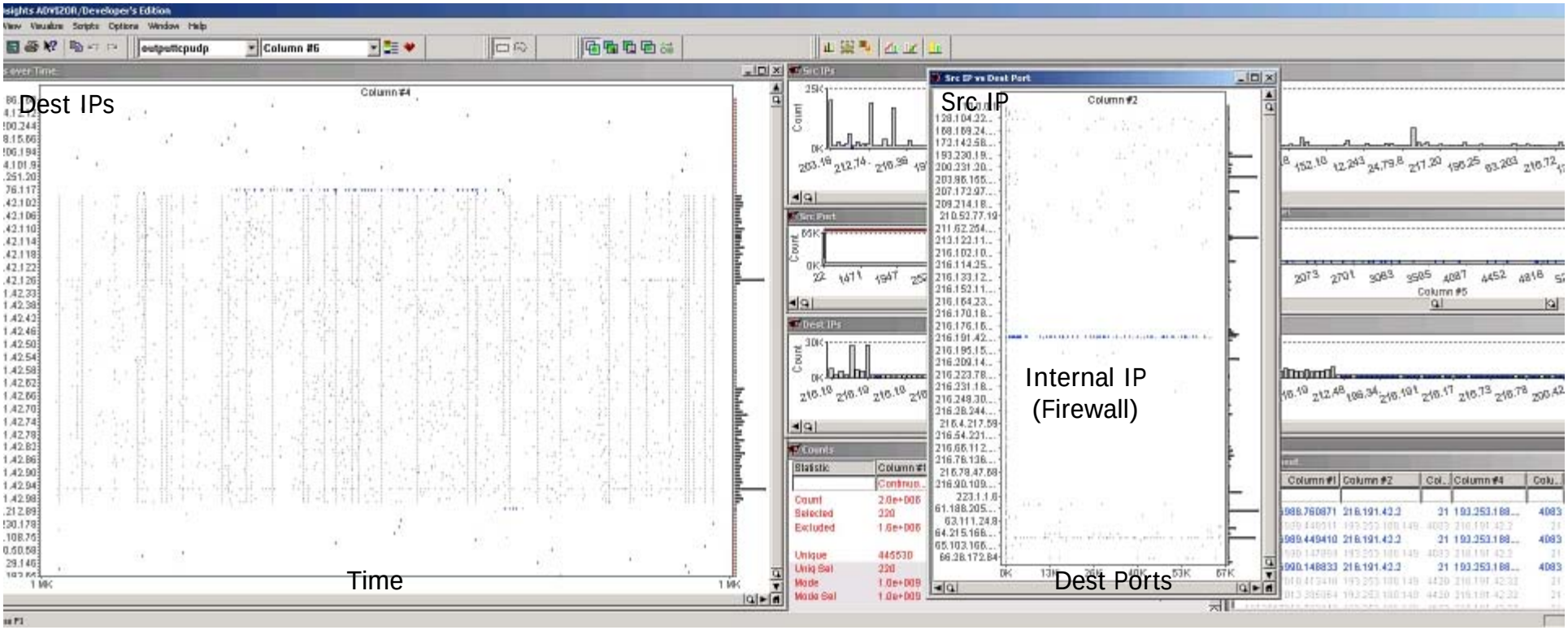
Confidential and © 2002 Oculus Info Inc. and Others

Tests

2M TCP and UDP Records

17

Slow Scan? one Src, many ports, over time.



On the weekend

Or Microsoft virus definition file update.

Confidential and © 2002 Oculus Info Inc. and Others

Animated Incident Explanation Engine

18

- Designed to show the significance and nature of the events without overwhelming the viewer
- Easy to see who did what to whom and when
- Excellent for explaining concepts to non-experts

- Analyst workbench identifies and isolates incidents. When an incident is isolated, all the records associated with it are written out and saved. Over time, a set of verified assessed incidents is created.
- The explanation engine works on this set of incidents.

Confidential and © 2002 Oculus Info Inc. and Others

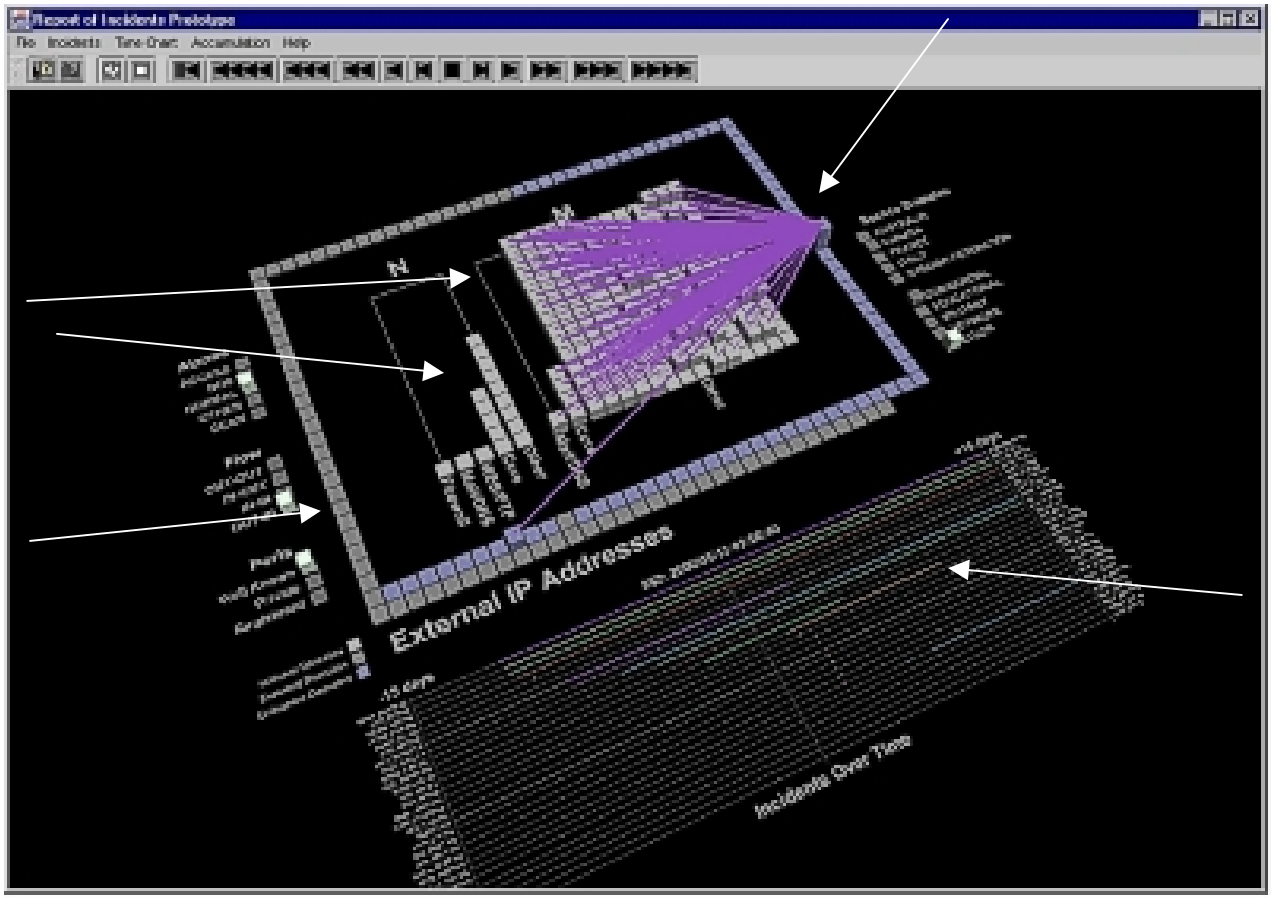
Animated Incident Explanation Engine

19

IP Half Scan is shown.

Internal IPs
-grouped by function

External IPs
- sorted by frequency
- and/or grouped by watchlist order



Timeline of Incidents

-one line per incident

Usage – show all records, show all records associated with one incident, animate over time all records, or animate records for one or more incidents. Playback speed is from fast fwd to very slow.

Confidential and © 2002 Oculus Info Inc. and Others

Future Developments

- Expansion and integration of the two current tools
- Anomaly detection capability through the use of network traffic data along with fused IDS alarms
- Integrated time based comparisons

Confidential and © 2002 Oculus Info Inc. and Others

Conclusions

- Visualization has proved to be an effective analyst's tool
- Massive amounts of information are easily understood by non-experts
- More development and research needed

Confidential and © 2002 Oculus Info Inc. and Others

Questions?

Confidential and © 2002 Oculus Info Inc. and Others